DOCUMENT RESUME

ED 372 099 TM 021 773

AUTHOR Uslick, JoAnn; Walker, Carole
TITLE An Evaluation of an Innovation: Standardized Test Scores Were Not Valid Indicators of Success.
PUB DATE Apr 94
NOTE 43p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1994).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Achievement; Accountability; *Achievement Tests; Beliefs; Cooperative Learning; Curriculum; Educational Change; *Educational Innovation; Educational Technology; Elementary Education; Mathematics; Program Evaluation; Research Methodology; *Scores; *Standardized Tests; Student Attitudes; Suburban Schools; Teacher Attitudes; Teacher Role; Urban Schools
IDENTIFIERS California Achievement Tests; *Lighthouse Learning Project

ABSTRACT

An overview is provided of the evaluation of the Lighthouse Project, an education enhancement project that began in one urban and two suburban districts. Its methodology was the context for showing how the focus on the results of a standardized achievement test in mathematics inhibited the implementation of mathematics reform in the elementary grades. It is evident that this teacher-driven innovation had a significant impact on teacher and student attitudes about mathematics and about technology as a tool for learning. Judgments based on the analysis of journal entries, surveys, interviews, observations, and California Achievement Test results support the following conclusions: (1) technology facilitates cooperative learning and individualized learning; (2) technology can be a catalyst for changing attitudes about mathematics and the teacher's role; (3) assessment of student learning should align with beliefs, curriculum, and instruction; and (4) accountability measures such as standardized tests should not be used to judge the impact of the program on student learning. Five appendixes, which contain seven figures, provide supplemental information about the evaluation and student achievement. (Contains 37 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 372 099

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
JoAnn Uslick
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

AN EVALUATION OF AN INNOVATION:

STANDARDIZED TEST SCORES WERE NOT VALID INDICATORS OF SUCCESS

JoAnn Uslick
Kent State University

Carole Walker
Coordinator of the Lighthouse Project
Tallmadge, Woodridge, and Akron, Ohio School Districts

A paper presented at the Annual Meeting of the American Educational Research Association.
New Orleans, Louisiana, April 4-8, 1994.

TM021773

BEST COPY AVAILABLE

ABSTRACT

By their nature, innovative educational programs that involve the use of technology have many interrelated variables of interest. But, understandably, stakeholders and funding agencies want the evaluation to focus on observable, measurable outcomes that specifically relate to the impact of the program on student learning. In this paper we give an overview of an evaluation, and its methodology, as the context for showing how the focus on the results of a standardized achievement test in mathematics inhibited the implementation of mathematics reform in the elementary grades. Because one of the project's goals was to serve as a model for other schools and districts, a variety of descriptive data were used. Other project goals were based on specific recommendations called for by the NCTM's *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989). The lessons learned each of the four years of implementation have guided the project to success in changing beliefs, curriculum, instruction, and assessment in mathematics. It became evident that this teacher-driven innovation had a significant impact on teacher and student attitudes about mathematics and technology as a tool for learning. The changes in attitudes have resulted in teachers, at many levels of implementation, becoming facilitators of learning who have moved from a dependency on textbooks and rote memorization of basic facts to the use of a problem-solving approach to mathematics in the context of cooperative learning and teacher networking. Basing judgements on the analyses of journal entries, surveys, interviews, observations, and the California Achievement Test results, the evaluation findings include: (1) technology facilitates cooperative learning and individualized learning; (2) technology can be seen as a catalyst for changing attitudes about mathematics and the teacher's role in the learning process; (3) assessment of student learning in mathematics should align with beliefs, curriculum, and instruction; and (4) accountability measures such as standardized tests should not be used to judge the impact of the program on student learning. Indicators of successful implementation of the Lighthouse Project relate to these findings and the criteria for effectiveness developed by the U.S. Department of Education's Program Effectiveness Panel (PEP). Specifically, teachers have initiated evaluation of appropriate tools for learning and assessment. The project that began in 2 suburban districts, was successfully replicated in an urban district during a time when a number of administrative and coordinator changes occurred. The implementation process has accelerated with the newer participants due to teacher networking and nine levels of project implementation have been identified. More important, teachers have formulated a list of beliefs for the innovative project that serendipitously align with the NCTM *Standards*. An implication for future evaluations of innovations is that tests that do not align with the goals of the program are invalid indicators of success because they send mixed messages to participants and inhibit them from fully implementing what is being called for in the reform effort.

AN EVALUATION OF AN INNOVATION: STANDARDIZED TEST RESULTS WERE NOT VALID INDICATORS OF SUCCESS

For the past four years, with financial help from a local foundation, the administrators of three public school districts have supported evaluators from a state university in the formative and summative evaluations of the impact of the innovative Lighthouse Education Enhancement Project (LEEP). The project is innovative because it is a collaborative effort among an urban school district and two suburban districts where 74 teachers in 6 elementary schools are making an impressive effort to implement recommendations called for in the National Council of Teachers of Mathematics' *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989). At the same time, the teachers are learning to use five classroom computers as part of a formative experiment attempting to implement local area network technology at the point of instruction in grades 1 through 5. Newman (1990) points out that "a formative experiment can involve elaborate arrangements for teacher training, curriculum development, and production of classroom materials to create an environment in which students and teachers can confront instructional tasks....Instead of rigidly controlling the treatments and observing differences in the outcome, as in a conventional experiment, formative experiments aim at a particular outcome and observe the process by which the goal is achieved" (p. 10). Thus, the evaluation of the Lighthouse project had to be decision-oriented and take a responsive, naturalistic approach using primarily ethnographic methods (Madaus, Haney, & Kreitzer, 1992). The purpose of this AERA paper presentation is to present a case for the use of a systemic approach in evaluating an innovation (Salomon, 1991), and to discourage the use of standardized tests results in evaluating the effectiveness of implementing the recommendations called for in the *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989).

Rationale

An understanding of the framework of this evaluation begins with knowing that the goals of the project were to improve elementary math teacher effectiveness and student competencies in critical thinking, cooperative learning, problem solving, and the use of technology. Another intended goal of the project was "to demonstrate that a technology enhanced mathematics curriculum will significantly improve student performance on standardized achievement tests." When the project began, this goal seemed reasonable even though the use of standardized achievement test scores as an outcome measure for evaluation of an educational program poses a potential threat to the validity of the evaluation because of the lack of alignment between the instructional program and the test content (Crocker, Llabre, & Miller, 1988).

The need for measurable outcomes is important for determining the worth of a program. Administrators and the public perceive that standardized tests are important for determining how the districts compare with other districts in the basic skills of mathematical computation and applications. But this perception has been shown by many critics of standardized tests to invalidate them due to the consequences of testing policies (Shepard, 1990; Herman & Golan, 1991; Nolen, Haladyna, & Haas, 1992; Gifford & O'Connor, 1992; Herman & Golan, 1993; Haney, Madaus, & Lyons; 1993). At the time of the commencement of the project a dilemma was evident because there were no valid and reliable alternative assessments available. Therefore, a norm-reference test became the only quantifiable evidence that could be used to show changes in students' achievement in mathematics. The standardized test that was available for students across the three Lighthouse public school districts was the California Achievement Test (CAT), Form E, using the 1984-85 norms. Whether or not these old norms would be a

problem in determining increased achievement was another issue to consider (Shepard, 1990). When the project began, the evaluators were not fully aware of the criticisms concerning the use of standardized tests. Consequently, the reasoning then was that what curricular and instructional changes in mathematics that took place due to the project should not adversely affect the basic computational and application skills in mathematics, and might even improve them. But the consequences of using the standardized tests was not considered.

The NCTM *Evaluation Standards* state that "the role of evaluation emerges as a critical component of reform....Many existing tests cannot measure the student outcomes identified in the *Standards*" (p. 189). That became very evident after the first year of implementation of the Lighthouse project. Teachers were in a quandary because they wanted to fully implement the goals of the program, but they also wanted their students to do well on the standardized tests that their district used. There were obvious conflicts between the two objectives. After four years of implementation of the project, many teacher participants believe that accountability measures are important, but these should not be the same as what is used to evaluate the impact of the program. Furthermore, teacher participants are requesting staff development on the use of alternative assessments such as portfolios, projects, and performances. One teacher's view of the problem illustrates the accommodations and frustrations of many teachers: "I will not teach to a test, however, I think we need an inservice on assessment in today's projects. The tests of the past, e.g., CAT, Iowa Basic, do not test like we teach." In a survey of teacher participants that was given during the fourth year of project implementation, 30 of the 36 teacher respondents communicated to evaluators that standardized tests, competency tests, and their report cards do not adequately evaluate the type of learning that was going on in the Lighthouse

classrooms (See Appendix A). This quandary will continue to inhibit implementation progress until valid measures are found or developed and pilot tested for practical use in the Lighthouse classrooms.

A mixed-method evaluation design was used during the four-year period of this study. Caracelli and Greene (1993) define mixed-method evaluation designs "as including at least one quantitative method (designed to collect numbers) and one qualitative method (designed to collect words), where neither type of method is inherently linked to a particular inquiry paradigm or philosophy" (p. 195). Thus, to understand the rationale for the use of mixed-methods in the evaluation of the Lighthouse Project, a consideration of the evaluation questions of the project is necessary. The evaluation questions were as follows: (1) What is the nature of changes in participants' knowledge, skills, and attitudes in the teaching of mathematics and technology's role in that process? (2) What is the nature of mathematics improvements for the students of the participating elementary teachers? (3) What is the nature of mathematics curricular changes in the participating school systems? (4) What impact will computers have on the teaching of mathematics if the computers are used as an instructional tool in the classrooms on a consistent basis? and (5) What is the nature of unanticipated outcomes of the Lighthouse Project? It is important to note from the above questions that there was an obvious need to study the changes of individuals within classroom environments that were changing as well.

In an effort to transcend the debate between the quantitative and qualitative research paradigms, Salomon (1991) effectively distinguished between the analytic and systemic approaches to educational research by giving two sets of studies as examples of the use of the two approaches. According to Salomon (1991),

> the systemic approach mainly assumes that elements are interdependent, inseparable, and even define each other in a transactional manner so that a change in one changes everything else and thus requires the study of patterns, not of single variables. It is, however, further argued that the validity of each approach is limited by the combination of assumptions made, phenomena chosen for study, questions asked, and research methodologies employed. Thus the two approaches, by epistemological necessity, have to be employed complementarily. (p. 10)

Clearly, the epistemological assumption that the three Lighthouse school districts, the six elementary schools, the 74 classrooms, and approximately 1800 students each year will interact and impact on each other indicated a need for a systemic approach rather than an analytical approach. But the variables that would be a part of the system were not readily identifiable and methodologies were not apparent.

In a discussion of computer supported collaborative learning (CSCL), Salomon (1992) contends that studying changes in individuals within a social context that is changing becomes

> a rather demanding task inasmuch as there are no well developed methodologies easily available to us. In fact, only recently, partly due to the increase in CSCL and partly due to the increasing dissatisfaction with so-called positivistic and reductionistic paradigms, has the study of individuals' change within a changing environment received serious attention (e.g., Altman, 1988,; Newman, 1990; Salomon, 1991). Clearly the analytic-experimental approach we are so familiar with cannot fully satisfy the need to study individual changes in a changing

context. (p. 65)

According to Salomon (1991), studying how one variable contributes to the outcomes of a project is "like asking how much did the flute, in a 120-piece orchestra, contribute to the quality of the music played" (p. 14). Salomon (1991, 1992) offered an appropriate methodology for the systemic study of classrooms using a special case of Multi-Dimensional Scaling (Guttman, 1969) called Smallest Space Analysis (SSA). But in the case of the Lighthouse Project, the nature of this formative experiment inhibited the anticipation of the extent of the impact of the project on participants and the development of appropriate measures to assess relevant variables before the project began in order to determine if there was a change in patterns. More important, budget constraints was a factor that contributed to limited instrumentation and personnel during the evaluation. CAT scores were available for all student participants and they were perceived by the public as important indicators of student achievement. Again, the use of in-place standardized tests seem to fulfill the minimum requirements for measuring outcomes of the project without calling on the expensive, time-consuming demands of developing other instruments and assessment protocols to determine if there was improvement of student learning in mathematics. But their convenience, cost efficiency, and limited definition of mathematics could not make up for the consequences of using them.

Although the Lighthouse goals are more substantive than the use of technology in the classroom, many participants continue to believe the computer has been the catalyst for change. According to Salomon (1992), "the computer may serve as a very useful subversive lever for change, but the change must encompass the whole learning environment....Its use shifts learning from recitation to exploration and construction, from being individually-based to being team-

based, and from being separated by disciplinary lines to being interdisciplinary" (p.63). These shifts in learning were consistent with the goals of the Lighthouse Project and, based on Salomon's orchestra metaphor, would be hypothesized to result in systemic change because of the dependency of variables on each other. Thus, in designing the methodology for the evaluation, the influence of technology on teachers and students had to be accounted for because of its catalytic qualities, in addition to representing the largest portion of a $1.5 millon budget and the biggest investment of teachers' time. It was also important to consider Salomon's (1992) argument that the effectiveness of technology depends upon the "orchestration of the whole learning environment---the curriculum, the activities that students engage in, students perceptions of the learning goals in the classroom, their social interactions, the teacher's behavior, and more" (p. 63). A need for an ethnographic study of the learning environments of a wide range of Lighthouse classrooms was apparent because teachers were free to implement the program as they felt comfortable with the recommendations. Participating in the culture of individual classrooms was the only way to describe the essence of <u>if</u>, <u>when</u>, and <u>how</u> changes would occur.

Another important consideration in designing the evaluation methodology involved data collection procedures and their relationship to project implementation. Apple (1992) addressed a number of crucial issues related to implementing the NCTM *Standards*. Of the five issues, the one most germane to the Lighthouse Project evaluation was "the complicated realities of teachers' lives." According to Apple (1992), intensification of teachers' work load, as called for in implementing the *Standards*, has lead them to "cut corners so that what is essential to the task immediately at hand is accomplished. ...Getting done is substituted for work well done. And as time itself becomes a scarce commodity, isolation grows, thereby reducing the possibility

of interaction and discussion among teachers to jointly share, critique, and rebuild their practices" (p. 426). The project required that teachers receive training in philosophy, use of math manipulatives, and the integration of technology and instruction; they were eventually asked to change the structure of their classrooms; and, finally, to share their successes and frustrations with other teacher participants. In other words, teachers had to fit more work into their crowded day, but at the same time teachers were be the key to successful implementation of the Lighthouse Project. How then must information be obtained from the key players of the innovation? The data collection and analyses procedures of the evaluation must allow for the understandable variability in teachers' cooperation and the lack of response to requests for information. Every effort possible was made to find out how teachers' beliefs and practice were changing, but at the same time not to add more to their workload. Follow-up procedures were used sparingly, realizing that teachers should attend to students', administrators', and the project coordinator's requests before our requests for evaluation information. Considering these accommodations, approximately 40% of the teachers cooperated with the major data collection procedures, although not always the same 40%.

After these basic considerations in evaluation methodology were resolved, the only unsettling one that remained was the adequate assessment of student learning in mathematics. NCTM President Mary Lindquist (1992) points out that there are four "legs" to consider when implementing the NCTM *Standards*. They include curriculum, instruction, assessment, and teacher beliefs. She contends, "A shift in any one of these four legs without a similar shift in the others will definitely leave us unbalanced. Thus, we must change assessment along with curriculum, instruction, and our beliefs as we move to empower every student in mathematics"

(p. 5). In a discussion of the NCTM *Standards* "revolution", Willis (1992) stressed, "standardized tests emphasize '19th century arithmetic skills,' even though math educators are united in the belief that students need to be able to do much more, especially problem solving. ...Because many tests haven't changed, and because parents are concerned about test scores, teachers are reluctant to shift the focus of instruction from what has traditionally been taught" (pp. 4-5).

Nolet and Tindal (1990) point out that valid interpretation of test scores is the shared responsibility of the test designer and test user. In their construct validity study of published achievement tests, of which the CAT test was one of the tests in the investigation, achievement test batteries were shown to be adequate measures of general achievement in the broadly defined construct of mathematics, but "inferences about student performance in skill areas represented by the various subtests included in most achievement batteries seem not to be supported" (p.2). It was concluded by the researchers that inferences based on the subtests are severely limited because "these tests fail to represent the wide range of classroom-relevant behaviors that are components of each construct...(and) can't provide information to support the inferences about the extent to which a particular curriculum works in a particular grade; the effectiveness of a particular teacher, or the outcome of a particular experimental intervention and they can't be ethically used for such purposes" (p. 22). The two parts of the CAT consist of a computation subtest and a concepts and application subtest. With the assumption that the project would impact achievement related to the content of second subtest and not the first, it was determined by the evaluators and district administrators that the subtest results should be looked at separately, even though Nolet and Tindal's construct validity study did not support doing this.

More support for not using standardized achievement tests was accumulating, along with the NCTM recommendations in the *Evaluation Standards* (1989) discouraging their use. An expansion of the evaluation standards can be found in the draft copy of NCTM's *Assessment Standards for School Mathematics* (1993b). It states:

> The weaknesses of standardized tests are many because they are often used as a basis for decisions that they were not designed to address. In particular, derived scores are invalid indicators of how much one knows. Also, aggregating standardized scores for students in a class (school, district, etc.) to produce a class profile of achievement (class mean) is both a very inefficient method of profiling and a meaningless indicator of achievement. The tests provide too little information in light of the cost involved. Unfortunately, their use appears to be more strongly related to political, rather than to educational, purposes....Finally, no claim of validity with respect to mathematical performance can be made. Standardized tests assume that mathematics is a single domain, rather than a collection of domains, and that all items reflect equivalent but independent concepts and procedure, rather than a network of structured, interdependent ideas. Scaling only involves counting the number of correct answers, not the reasoning or the strategies used to find the answer. (pp. 222-223)

The considerations involving the use of standardized tests that were mentioned above became obvious in the findings of the evaluation and inspired the goals of this paper. We will attempt to briefly describe here the magnitude of the project and the understanding and accommodations that were necessary as participants' beliefs, curriculum and instruction changed

while accountability measures of assessment stayed the same. The consequences of using a standardized achievement test to make judgements about the Lighthouse Project will be considered. Thus, the focus of this paper will be to give an overview of the evaluation of an innovation and to make a judgement about the validity of the CAT test as a component of the Lighthouse evaluation in order to determine its usefulness in future evaluations.

Evaluation Methodology, Data Sources, and Analyses

A mixed-method design was used throughout the four years of the Lighthouse Project and required the collection of data from administrators, teachers, and students who participated in the project. Very little quantitative data were available that directly related to project outcomes, other than the standardized test scores that were mentioned previously. Teachers were willing to use alternative assessments, but their availability, cost, and the need for pilot testing were issues to be resolved. The use of only one quantitative measure would seriously impact the integrity of a mixed-method design because the design implies the use of both quantitative and qualitative data. With the overabundance of qualitative data and only one source of quantitative data, it overemphasized the importance of the standardized tests.

For the purposes of a summative evaluation at the end of the fourth year of implementation, qualitative data were transformed into quantitative data so that the levels of implementation of each teacher participant could be determined. A rubric originally developed by Hord and Hall (1986) was adapted by the evaluator and the project coordinator for the purpose of defining nine levels of Lighthouse Project implementation and informing the stakeholders of progress towards implementation. (See Appendix B.) A mini-case study was used to illustrate the early frustrations of teacher implementation of the project and to gain

insight on teacher non-response to requests for data. Interview transcripts, observation notes, documents, and video-tapes were used to study and describe the implementation process relating to teachers and students. Open-ended surveys were used to validate information and as a follow-up for obtaining information related to beliefs and perceptions about technology and the assessment of student learning. Unstructured journal entries of teacher participants were the primary data source for studying the unexpected outcomes of the project. Entries were qualitatively analyzed and emergent domains identified.

Qualitative and quantitative data were integrated for the purposes of initiation, development, triangulation, complementarity, and expansion. Greene, Caracelli, & Graham (1989) cited these five purposes for mixed-method evaluations and grounded them both in theory and practice. All five purposes were relevant for integrating the qualitative and quantitative data for summary purposes at the end of four years of implementation. The four analytical strategies recommended by Caracelli and Greene (1993) were also valuable in "making sense of" these data. These recommendations include: data transformation, typology development, extreme case analysis, and data consolidation/merging. The fact that many purposes were being accommodated and a variety of strategies were needed indicates the impact of each variable on other variables and the importance of studying the system of variables that go to make up an innovative program. The implication was the need to study the change in patterns rather than the changes in isolated variables of interest (e.g., achievement in mathematics, levels of implementation, teachers' beliefs, attitudes towards technology, reasons for non-response, etc.). But when the project began, the study of the system of variables was not possible because the variables that made up the system had not yet been identified. It is now possible to put together

a system of variables and to use Salomon's (1991) recommended methodology for focusing on "complex patterns and changes thereof." (See Appendix C for an illustration of possible systemic study.)

Conclusions

Teachers were given the philosophical and practical training, the essential tools, and the administrative support necessary to implement recommended changes in curriculum and instruction called for by NCTM. It is important to note here that the teachers were under no pressure or time restriction to initiate change. The result of this was that the teachers had a lot of input into implementation procedures and the need for further training.

There was an immense amount of anecdotal evidence that this teacher-driven innovation has had a significant impact on administrators', teachers' and students' attitudes about mathematics and technology. Some of the reasons given by teachers for believing that the goals of the project could not have been accomplished without technology include: (1) with computers math makes sense, math has meaning, math is fun; (2) the computer reinforces the understanding of math concepts; (3) technology makes us aware that change in attitudes towards mathematics is necessary; (4) technology provides both individual and small group emphasis; (5) technology is multisensory where the manipulative activities are not; (6) the computer is there to respond immediately to errors as well as successes; (7) children develop self-esteem by using technology and helping others; and (8) computers increase teachers' enthusiasm. The technology component of the project made one teacher realize "my new and different role as a teacher. My ability to make a child think. --- And to find and create those situations for that, or those children, has added life to teaching. I love having the freedom to find new and different methods of teaching.

--- What an exciting dimension!" Another teacher wrote, "Not only are children reinforcing their skills in a nonthreatening way, they are also able to fuse knowledge, apply skills learned, and experiment and make discoveries."

Teachers also feel that cooperative learning is facilitated using computers. A teacher was asked to explain how; she wrote: "Just walk through any classroom that has more than one child on a computer and your question will be answered. There is no doubt that two minds are better than one when you see this happening. God gave children such inquisitive minds which they use for asking incredible questions, and for justifying answers. There is no need to justify an answer if there is no one to question it. Peers certainly question!" About asking children to work together at the computer, another teacher wrote: "Students tend to share their expertise with others. More often than not, the low achiever seems to be more computer literate than the academically successful student and they find common ground to discuss what they are doing on the computer." Finally, many teachers expressed that they were learning along with their students. This new role of a teacher learning from students and the power of group learning was expressed by a teacher as follows:

> I'm enthusiastic about the academic and social benefits which result from cooperative math groups. Students are aware of the power of "the group" to solve problems. Certainly the computer enhances cooperative learning. Since I'm learning too, the students help me learn and grow as well.

It is significant that these statements were made by elementary teachers who only recently were exposed to technology and the constructivist theories about knowledge and learning.

Teachers feel that math is now fun to teach. Consequently, they spend more time with

their students questioning, asking for reasons, looking for patterns, and discovering conceptual understandings. Their students are benefitting by learning to think mathematically. Teacher participants, at a variety of levels and stages of implementation, have become facilitators of learning and have moved from a dependency on textbooks, rote memory, and worksheets to the use of a problem-solving approach to mathematics in the context of cooperative student learning and teacher networking. The lessons learned over the past four years have guided the project to successful implementation of changes in beliefs, curriculum, instruction, and alternative assessment of elementary mathematics, in addition to the current expansion from the math curriculum to other subjects. As a natural, teacher-initiated consequence, the project is now moving into the upper grades and is continuing to serve as an innovation model for other school districts. (Uslick, Anglin, Jones, Brewer, & Shapiro, 1991; Uslick, Gill, & Godin, 1992; Uslick & Gill, 1993)

Criteria for judging exemplary programs have been developed by the Department of Education's Program Effectiveness Panel (PEP) and programs are judged accordingly in an effort to disseminate findings and promote their replication throughout the nation (Ralph & Dwyer, 1988). Three general questions addressed by the PEP effectiveness criteria include: (1) Is the evaluation credible? (2) Are the results of the program meaningful? and (3) Does the object of evaluation (the program or product) have the potential to be replicated? In order to answer the first question, an evaluation must consider the following:

> employ appropriate measurement (that is, instruments that are in line with the program's goals), technically strong measurement techniques, and careful, well-documented data collection. In addition, the evaluation must be able to link

obtained outcomes with the program itself. In other words, alternate explanations of results must be addressed and ruled out. Further evidence of a credible evaluation design is provided by a comparison standard, usually a carefully composed control group or appropriate norm groups. (Madaus, et. al, 1992, pp. 24-25)

The fact that the CAT was the only mathematics achievement test available among the three districts seriously affected the credibility of the evaluation. The criteria used by PEP also requires the use of a control group. This was attempted at the onset of the Lighthouse Project, but other districts that were approached would not cooperate with the evaluation procedures (e.g., making standardized test scores available, teachers keeping journals, allowing evaluators to observe, etc.). In addressing the other two questions, the meaningfulness of the program was evident in the fact that there was a state and national reform movement occurring in how and what mathematics were taught and districts' needs assessment surveys showed that change was being called for by the districts' administrators and teachers in the elementary grades. The replicability of the program was evident because the program started in three suburban schools during the first and second years of the program; then it was successfully duplicated in three urban schools during the third and fourth years of the program. Another important issue to consider is that there were two coordinators of the project during the four years of implementation, and changes in three administrators. Many times when key people are replaced, innovative projects do not continue with the success as was first experienced. On the contrary, the Lighthouse implementation process was accelerated when new people became involved.

Other relevant information dealing with evaluating an innovative program has to do with

the evaluation questions and the claims made about (1) the academic achievement of students, i.e., changes in mathematical knowledge and skills; (2) improvements in teachers' and students' attitudes and behaviors; and (3) improvements in instructional practices and procedures (Madaus, et.al., 1992). Related to the second and third of these claims, the analysis strategies used with the qualitative data, and the transformation of these data to produce quantitative data, have shown that, indeed, teachers' attitudes, behaviors, and instructional practices have improved. Students' attitudes and behaviors towards mathematics have also improved. (See Appendix D for selected evidence, e.g., a summary of an interview with a non-responding teacher and two administrators' reactions towards the Lighthouse Project.) The first of these claims that relates to academic achievement would infer a question about the statistical significance of changes in CAT scores over the length of the evaluation period. The results show that statistical significance was evident and showed improvement only in the school district that began the project in kindergarten. For the district that began implementing the project in 3rd grade, the scores were significantly lower from 2nd- to 4th-grade, but the scores improved by 6th grade. For the district that had only one year of implementation, the scores did not change significantly for two of the three schools. (See Appendix E for CAT results by district.) But it is the consensus of the evaluators, the teachers, and the administrators who are in touch with the classrooms that the standardized tests currently used by the three districts have not been relevant to project goals. It is now obvious that the CAT test results gave little or no information to stakeholders about the type of curriculum and evaluation that is recommended in the NCTM *Standards*.

Therefore, during the four years of this evaluation study, it has become apparent that the

goal related to an increase of standardized test scores was not appropriate in the context of this project. Other researchers agree. According to Romberg, Wilson, Khaketla, & Chavarria (1992), "A major argument against standardized tests has been their failure to assess higher-order skills; rather, such tests emphasize computations, recognition, and other lower-order thinking skills (Meir, 1989; Putnam, Lamper, & Peterson, 1989)" (p. 63). Romberg & Wilson (1992) have studied the alignment of standardized tests with the NCTM *Standards* and have found "the currently used standard tests at grade 8 are not valid instruments for assessing the content, processes, and levels of knowledge called for in the *Curriculum and Evaluation Standards*" (p. 22). And according to a three-year NSF-supported study by the Center for the Study of Testing, Evaluation, and Educational Policy (CSTEEP); "only 3 percent of the questions on standardized mathematics exams tests conceptual knowledge and only 5 percent test for problem-solving and reasoning skills" (NCTM, 1993a, p. 7).

It is important to note that standardized achievement tests do not depend upon setting educational standards, as is often assumed by the public. The test scores are norm-referenced, which means that a student's score is obtained by comparing the student's raw score to a relative standard, i.e., the norms of some defined reference group. Webb (1992) contends, "since the reference is to some characteristic of a group, a score on a mathematics norm-referenced test does not define what mathematics a student knows or does not know" (p. 675). The NCTM *Standards* are very clear about what mathematics students should know in the elementary grades, the middle-school grades, and high school. The word "mathematics" being plural indicates the multiplicity of components. The *Standards* emphasize learning processes through conceptual understanding; mathematical reasoning; connections among concepts, procedures and topics; and

problem solving as the context for learning. Thus, to evaluate the impact of the Lighthouse Project it is meaningless to use a standardized achievement test that is norm-referenced and was developed ten years ago during a time when computation dominated the mathematics curriculum. Moreover, since hands-on learning with math manipulatives and technology is used by children in Lighthouse classrooms to construct their knowledge and to make connections, valid assessments of learning in mathematics cannot be made without the use of manipulatives.

Importance of Study

From the previous conclusions, variables that will be used with the systemic approach have begun to be identified. These should help with future evaluations of the project, in addition to helping other evaluators of innovative programs that are trying to implement NCTM's recommendations. Alternative assessments are becoming available that align with the *Standards*. One, in particular, has been identified by the Lighthouse Project as an exemplary prototype. *Measuring Up: Prototypes for Mathematics Assessment* (1993), a report of the Mathematical Sciences Education Board and the National Research Council, is a collection of 13 tasks for fourth-grade students that "are intended to illustrate possible directions for new assessment instruments,...for children who have had the full benefit of a *Standards*-caliber mathematical education in kindergarten through fourth grade" (p. 7). Fortunately, the teachers in the Lighthouse Project are willing to look at these tasks and to seriously consider them for use, as they were meant to be looked at and used. That is, they "illustrate directions for tomorrow"; "set targets for teaching and learning"; and to help define "appropriate goals for fourth-grade instruction." More important, students in the Lighthouse Project are fortunate to have achieved varying degrees of "a *Standards*-caliber mathematical education."

Knowledge comes from a combination of data and assumptions. In specifying the direction of this evaluation study, the stakeholders have assumed that if teachers believe in the NCTM *Standards* and teach accordingly, the students will be empowered to think mathematically. The assumptions and data were closely considered and the quantitative data were found lacking with respect to evidence that showed that students were thinking mathematically. The significant finding of this evaluation is that currently used standardized test results for mathematics achievement are not valid indicators of success of the implementation of the NCTM *Standards*. The multi-faceted spirit of the *Standards* demands much more than knowledge recall and procedural skill. Therefore, evaluators need to consider more than the results of standardized tests when assessing the impact of innovative programs because the consequences of not doing so might impact the implementation process. In the case of the Lighthouse teacher participants, many were reluctant to implement the project because of being judged by students' performance on the CAT test. In sum, in order to address the consequential aspects of validity called for by measurement specialists (Messick, 1989; Moss, 1992; Shephard, 1993) and the latest recommendations for validating exemplary programs (Walberg & Niemiec, 1993), evaluators need to be explicit about how the use of standardized tests affects project implementation and to support the drive toward the development and validation of alternative assessments of student learning.

REFERENCES

Altman, I. (1988). Process, transactional/contextual, and outcome research: An alternative to the traditional distinction between basic and applied research. Social Behavior. 3, 259-280.

Apple, M.W. (1992). Do the Standards go far enough? Power, policy, and practice in mathematics education. Journal for Research in Mathematics Education. 23(5), 412-431.

Caracelli, V.J. & Greene, J.C. (1993). Data analysis strategies for mixed-method evaluation designs. Educational Evaluation and Policy Analysis. 15(2), 195-207.

Crocker, L., Llabre, M., & Miller, M.D. (1988). The generalizability of content validity ratings. Journal of Educational Measurement. 25(4), 287-299.

Gifford, B.R. & O'Connor, M. C. (Ed.) (1992). Changing assessments: Alternative Views of Aptitude, Achievement, and Instruction. Norwell, MA: Kluwer Academic Publishers.

Greene, J.C., Caracelli, V.J., & Graham, W.F. (1989). Toward a conceptual framework for mixed-method evaluation designs. Educational Evaluation and Policy Analysis, 11, 255-274.

Guttman, L. (1969). A general nonmetric technique for finding smallest coordinate space for a configuration of points. Psychometrics, 33, 465-506.

Haney, W.M., Madaus, G.F., & Lyons, R. (Ed.) (1993). The Fractured Marketplace for Standardized Testing. Boston: Kluwer Academic Publishers.

Herman, J.L., & Golan, S. (1991). The effects of standardized tests on teachers and learning --- Another look. CSE Technical Report 334. CRESST/UCLA Graduate School of Education, Los Angeles, California.

Herman, J.L., & Golan, S. (1993). The effects of standardized tests on teaching and schools. Educational Measurement: Issues and Practice, 12,(4), 20-25, 41-42.

Hord, S.M. & Hall, G.E. (1986, April). Institutionalization of innovations: Knowing when you have it and when you don't. In Murphy, S.C.; and others, School improvement: Messages from five years of research. Symposium presented at the annual meeting of the American Education Research Association, San Francisco.

Lindquist, M.M. (1992). Assessment: Why? Because! NCTM News Bulletin, November issue, p. 3 & p. 5.

Madaus, G.F., Haney, W. & Kreitzer, A. (1992). Testing and Evaluation: Learning from the projects we fund. New York: Council for Aid to Education.

Mathematical Sciences Education Board & National Research Council. (1993). Measuring Up: Prototypes for Mathematics Assessment. Washington, DC: National Academy Press.

Meir, T. (1989,Jan/Feb). The case against standardized achievement tests. In Rethinking Schools, 3,(2), 9-12.

Messick, S. (1989). Validity. In R.L. Linn (Ed.), Educational measurement (3rd ed., pp. 13-103). Washington, DC: American Council on Education and National Council on Measurement in Education.

Moss, P.A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. Review of Educational Research, 62,(3), 229-258.

National Council of Teachers of Mathematics. (1989). Curriculum and evaluation standards for school mathematics. Reston: Author.

National Council of Teachers of Mathematics. (1993a). Report claims standardized tests do not make the grade. NCTM News Bulletin, January, 1993, p.7.

National Council of Teachers of Mathematics. (1993b). Assessment standards for school mathematics. Reston: Author.

Newman, D. (1990). Opportunities for research on the organizational impact of school computers. Educational Researcher, 19(3), 8-13.

Nolen, S.B., Haladyna, T.M., & Haas, N.S. (1992). Uses and abuses of achievement test scores. Educational Measurement: Issues and Practice, 11,(2), 9-15.

Nolet, V. and Tindal, G. (1990, April). Evidence of construct validity in published achievement tests. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA.

Putnam, R.T., Lampert, M., & Peterson, P.L., (1989). Alternative perspectives on knowing mathematics in elementary schools. Center for the learning and teaching of elementary school subjects. Michigan State University, Lansing.

Ralph, J., & Dwyer, C.M. (1988). Making the case: Evidence of program effectiveness in schools and classrooms -- Criteria and guidelines for the US Department of Education's Program Effectiveness Panel. Washington, DC: US Government Printing Office.

Romberg, T.A. & Wilson, L.D. (1992). Alignment of tests with the standards. Arithmetic Teacher, 40(1), 18-22.

Romberg, T.A., Wilson, L.D., Khaketla, M., & Chavarria, S. (1992). Curriculum and test alignment. In Romberg, T.A. (Ed.) Mathematics Assessment and Evaluation: Imperatives for Mathematics Educators. Albany: State University of New York Press.

Salomon, G. (1991). Transcending the qualitative/quantitative debate: The analytic and systemic approaches to educational research. Educational Researcher, 20(6), 10-18.

Salomon, G. (1992). What does the design of effective CSCL require and how do we study its effect? SIGCUE Outlook, 21(3), 62-68.

Shepard, L.A. (1990). Inflated test score gains: Is the problem old norms or teaching the test? Educational Measurement: Issues and Practice, 9,(3), 15-22.

Shepard, L.A. (1993). Evaluating test validity. In L. Darling-Hammond (Ed.), Review of research in education (Volume 19). Washington, DC: American Educational Research Association.

Uslick, J., Anglin, L., Jones, J., Brewer, R., & Shapiro, W. (1991). Lighthouse Education Enhancement Project: Year-end Assessment Report. Bureau of Research Training and Services, Kent State University, Kent, Ohio.

Uslick, J., Gill, N., & Godin, K. (1992). Lighthouse Education Enhancement Project: Year-end Assessment Report. Bureau of Research Training and Services, Kent State University, Kent, Ohio.

Uslick, J. & Gill, N. (1993). Lighthouse Education Enhancement Project: Year-end Assessment Report. Bureau of Research Training and Services, Kent State University, Kent, Ohio.

Walberg, H.J. & Niemiec, R.P. (1993). Validating exemplary programs: Methods and criteria. Educational Evaluation and Policy Analysis, 15,(4), 429-436.

Webb, N.L. (1992). Assessment of students' knowledge of mathematics: Steps toward a theory. In Grouws, D.A. (Ed.), Handbook of Research on Mathematics Teaching and Learning. New York: Macmillan Publishing.

Willis, S. (1992). Mathematics education: Standards 'revolution' takes hold. Curriculum Update of ASCD, January, 1992, pp. 1-6.

APPENDIX A

Eighty-three percent, or 30 out of 36 teachers who responded to the survey, felt that standardized tests do not adequately evaluate the type of learning that goes on in a Lighthouse classroom.

Question: **DO YOU FEEL THAT STANDARDIZED TESTS ADEQUATELY EVALUATE THE TYPE OF LEARNING THAT GOES ON IN A LIGHTHOUSE CLASSROOM?**

No, the students may know the material but cannot handle the strange format.

Standardized tests do not have enough of any one type of problem to adequately assess standards.

They are adequate for 2/3 of my class. (For) the other 1/3, they are not, as these students need more of "hands-on" evaluation measures.

No, because manipulation is not being tested. Neither is the process.

There is no one test that adequately evaluates the "whole" learner. I would much prefer to see a portfolio type assessment used.

All of the evaluation tools should be looked at and changed to fit the way in which the children are learning.

No. Much of what we do in Lighthouse is hands-on. This requires observations and dialogue with each student to measure understanding.

No. The standardized tests do not necessarily cover what has been taught.

None of the evaluation tools reflect the hands-on emphasis that is a major component of the Lighthouse project.

Some (students) are having a good day and some a bad day, so these scores aren't always accurate.

We are all struggling with a form of evaluation which tells us what to tell parents about what their child is capable of and is doing (performance) in class day to day.

Standardized tests and proficiency tests do not adequately evaluate the type of learning that goes on in a Lighthouse classroom. These tests are not able to measure the process by which the student solves mathematical questions.

I don't feel that standardized tests test the way we teach. They are usually multiple choice and leave little chance for extended thinking. In my classroom, I try to observe many of the outcomes based on student performance with technology and manipulative activities. I struggle with organizing my records and getting an accurate evaluation for each student.

No. Very little room is given for problem solving, manipulative use, or computer use. The standards we use for reporting are designed for our "old" style of teaching.

No I do not. Some type of a check evaluation is needed. We need to evaluate what they can do and can't do. The CAT was a disgrace in that it only had one or two samplings of some skills, e.g., money, telling time.

No. No use of manipulatives that we allow for.

None of the tests adequately evaluates the use of manipulatives in learning math skills.

Standardized tests tell you very little of what a child can do.

I don't think standardized tests in any of the core areas (math and reading) assess what or how we teach.

Standardized tests only test computational skills, not "how" or the thinking or reasoning skills.

A standardized test has no relationship to the use of the computer.

No, very little problem solving assessment.

In order to assess students, many things must be considered including all the different methods of learning and instruction which takes place. To have a child complete standardized tests or proficiency tests puts the weight of all the learning in one basket. These tests don't allow students to show all they really know, only select concepts they do or don't know.

No. I'm anxious to see how they do on the CAT next fall.

No. Very little problem solving assessment.

No. CAT test did not focus on "why we do what we do" in math.

The usual paper-pencil tests of basic skills are not sufficient to determine a student's problem solving ability. To be effective, assessment should employ a variety of methods. A broadbased assessment can give a valid picture of a students problem-solving skills.

No. The tests are not presented in the same way that the children have been taught.

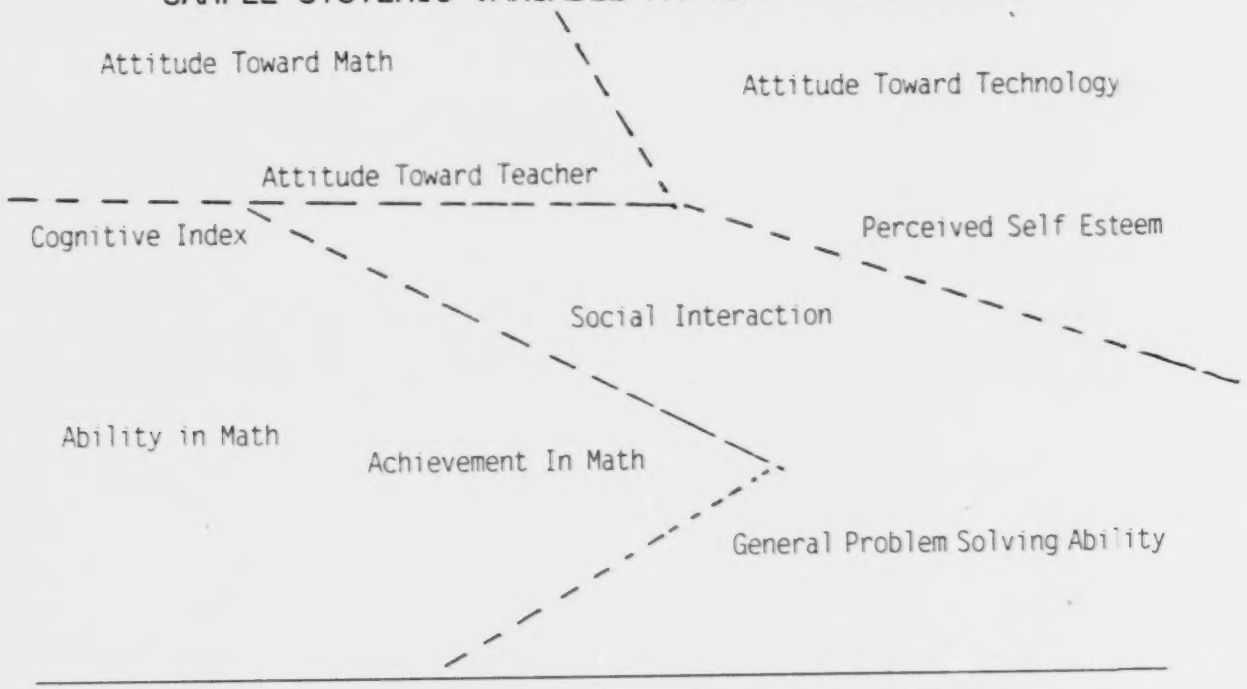
APPENDIX B

LEVELS OF USE OF LIGHTHOUSE PROJECT IMPLEMENTATION

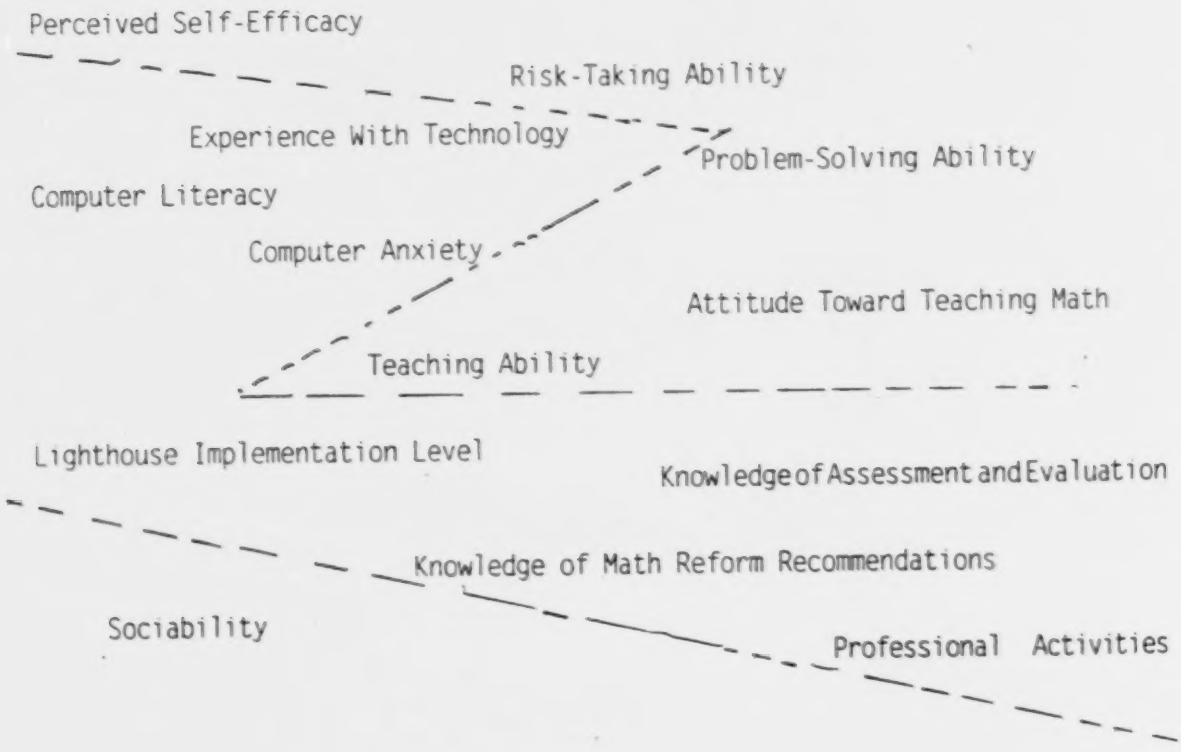
Level of Use		Behavioral Indices of Level
0	Nonuse	No observable change. No action is being taken with respect to the innovation.
1	Orientation	The user is seeking out information about the innovation.
2	Preparation	The user is preparing to use the methods and tools of the project.
3	Mechanical Use	The user has little understanding of the changes in teaching and learning.
4	Routine	The user is making a few changes. Occasional use of manipulatives and computers. Predominantly uses a traditional manner of teaching.
5	Refinement	The user is making changes to increase student outcomes. Demonstrates appropriate use of manipulatives and computers. Limited problem-solving activities.
6	Awareness	User concentrates on higher levels of thinking, encourages problem-solving, uses computer creatively.
7	Integration	The user is making deliberate efforts to coordinate with others in using innovation.
8	Renewal	The user is seeking more effective alternatives to the established use of the innovation.

APPENDIX C

SAMPLE SYSTEMIC VARIABLE PATTERN FOR STUDENTS



SAMPLE SYSTEMIC VARIABLE PATTERN FOR TEACHERS



APPENDIX D

An Interview with a Non-Respondent

During the last week of May, 1993, the evaluator interviewed a Phase III Lighthouse participant who had been teaching for many years. At the time of the interview, she had not responded to any of the three surveys nor had she submitted a journal for either of two collections during 1992-93. The coordinator had asked the evaluator to interview this teacher because, over the past few months, the coordinator had noticed a dramatic change in the teacher's attitude toward Lighthouse. Although teacher interviews were not part of the planned data collection this year, the evaluator agreed to this interview because it offered possible insight into one teacher's non-response to data collection requests.

During the last week of May, 1993, the evaluator set an appointment to meet with the teacher after school in her classroom. The interview was to be guided by the questions that had been used for teachers' interviews in previous years. Interviews with teachers typically lasted 15-30 minutes and were usually audio taped. However, this teacher declined to be audio taped during the interview. The interview lasted two hours. At the end of that time the teacher said, "I guess I interviewed you more than you interviewed me." The evaluator felt her time was well spent in the interview. She now knew that the reason for this teacher's lack of response to data collection was her uncertainty about what Lighthouse was about and her belief that she wasn't qualified to give feedback for assessment purposes. She was just beginning to understand the changes called for in the NCTM Standards and had many questions regarding the need for change and specific ways change could be implemented. Toward the end of the interview she asked if she could obtain a copy of the NCTM Standards document. The evaluator suggested asking the coordinator to copy specific pages relating to the primary grades.

Seven basic points emerged in the interview:

1. Although she had received training in Math Their Way a few years before, this teacher did not consider herself a part of Lighthouse until she started using her five classroom computers.
2. She highly valued the usefulness of basic facts in math education but was beginning to reevaluate that priority because there were so many other topics to discover.
3. She wanted to delve deeply into mathematical topics and was frustrated that she did not have the time to do so.
4. She never really enjoyed teaching math until this year.
5. She taught without a math textbook for the first time this year. When asked why, she said that her Math Their Way training, combined with the technology, had forced her to reevaluate the textbook material. She now felt that some of it was irrelevant and confusing and classroom time was better

spent using hands-on material.

6. She had to overcome a dislike and fear of technology based partly on the inconvenience of having to previously share one computer between several classrooms. She didn't like the "fuss"of pushing a cart, dealing with disks, etc.

7. She credits the Lighthouse coordinator and her own students with helping her overcome her fear of technology and see its advantages for teaching and learning mathematics.

In response to this teacher's expressed lack of understanding about the project, its goals, and the use of survey and journal data, the evaluator told the teacher she was not alone in her inexperience, dilemmas, and frustrations. The evaluator then stressed the importance of teacher feedback so these issues could be addressed by people who could offer assistance and support to her and others like her. The teacher then quickly read the two surveys, began to find meaning in them, and became interested in contributing to the assessment of the Lighthouse project.

Finally, the evaluator and the teacher discussed assessment of student learning. The teacher said competency tests really bothered her. She said her children had no problem passing them and she feared they were getting the wrong message about the tests' purpose. She was concerned parents and students aren't realizing that only minimum competencies are being assessed. She feels the tests should be harder so students don't think minimum competencies are all that is being asked of them. She was also concerned that competency items are written separately by individual districts and varied greatly from district to district. She related a recent problem she had with two new students from a neighboring district who had no problem with the competency test but could not keep up with her classroom lessons. She was uncertain how to explain the situation to the students and their parents. She felt she could use some in-service workshops on alternative methods for student assessment.

Two Principals' Reactions to Lighthouse

An open invitation was made to administrators to provide statements that would be included in this report regarding their reactions to the project. Two principals submitted the following comments.

As principal of a large urban school which houses the Lighthouse project in grades one and two, I would like to offer these comments regarding the program thus far:

TEACHERS TRAINED IN THE PROJECT

1. Varied in terms of math philosophies yet subscribed to the NCTM Standards.
2. Varied in terms of computer knowledge and use.
3. Were all receptive to the project and believed the technology piece would provide varied experiences for youngsters all along the mathematics curriculum continuum.

PARENTS OF THE STUDENTS IN THE PROJECT

1. Were and continue to be invited to participate by interacting with their child/children and the computer.
2. Are impressed with the mathematics knowledge and vocabulary usage their children possess.
3. Would benefit from some training in the NCTM Standards, how the computers fit in and their place (support) in the project.

STUDENTS IN THE PROJECT

1. Like to use the computer!
2. Are developing basic computer skills rapidly as a result of use.
3. Are able to merge programs.
4. For students who find fine motor skills troublesome because of the time it takes to form letters, etc., the computer allows them the ability to get ideas, concepts, manipulations completed quickly thus allowing more time on content.
5. See and hear immediate feedback and have opportunity to self-correct.
6. Have the power of creating text and controlling content.

The principal ended the statement by saying, "It is an exhilarating project, worthy of continuation through the grades."

Another principal of an urban elementary school commented directly on the questions in the open-ended teachers' survey. This principal's comments regarding the importance of technology to the project follow.

> I feel the goals of the project can be accomplished without the technology but I feel the technology has enhanced the project.
>
> The technology has created a new mindset for the teachers. They have changed their approach to math and their attitudes because of adapting the technology as part of the teaching process in the teaching of mathematics.
>
> The students attitudes have changed because we have taken advantage of the natural curiosity of a child and centered the learning on a discovery process.

Regarding the use of technology in facilitating individualized learning:

> I think computers have the capability of facilitating individualized learning very well. The students receive immediate feedback on their responses, students are able to work through a program at their own pace, the reteaching takes place when they need it, and the learning can follow a sequential pattern where prerequisite skills are mastered before moving onto the next challenge.

Regarding the use of technology in facilitating cooperative learning:

> Computers can facilitate cooperative learning to some extent, however I feel most of the interaction occurs between computer and student. The students do help each other when they have a problem. There is more cooperative learning taking place when they work with partners, but not as much as when they are working with manipulatives.

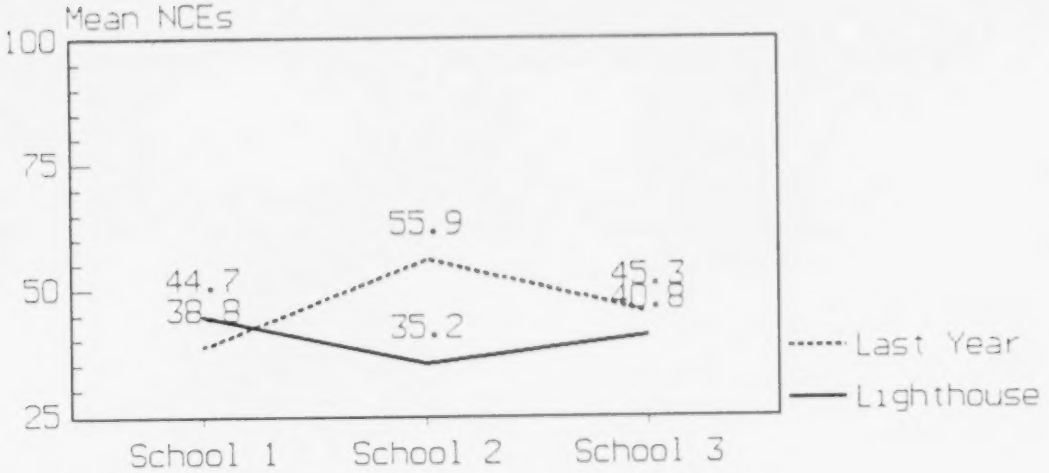
The principal's beliefs about assessment of student learning follow.

> Student learning can not be totally evaluated by tests that give an indication of learning at one moment in time. A more comprehensive form of assessment needs to be looked at to get a true indication of the learning taking place.
>
> Standardized tests have poor items at times, some students are test phobic and do not test well, and there is also a personal factor that may affect test scores. Report cards are ranking systems, giving a student his/her rank in comparison to their classmates. This does not reflect the individual learners' growth. Proficiency tests have the same downfalls as standardized testing.
>
> We are in the process of adopting a nongraded report card for Grades 1 and 2. We feel this is a step in the right direction. We are also talking about portfolio assessment and developing a skills checklist to assist in the evaluation procedures.
>
> The teachers tend to still use many of the tests that have been available, however I feel they also weigh personal observation more when assigning a grade. I would like to see the tests be revised to reflect a more problem solving approach.
>
> Assessment should drive instruction and curriculum. Teachers need to be trained to pre-assess students so the curriculum and instruction is tailored to the needs of the students. Why teach a concept if the students have already mastered it? Teachers should constantly monitor the students' behaviors and adjust the instruction and curriculum accordingly.

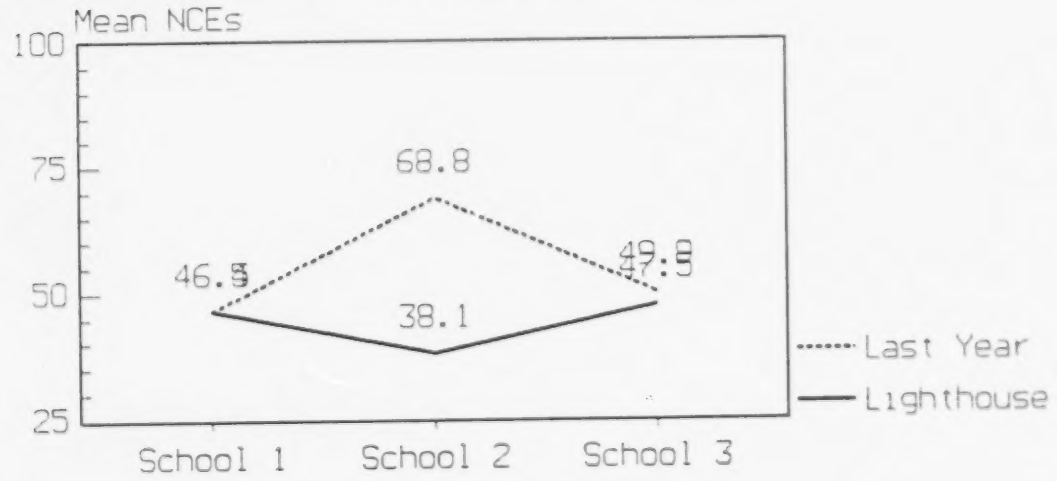
The principal's letter concluded with "I hope my comments are helpful to your project." Indeed they are, because it is obvious that the principal has taken a personal interest in the project and his reactions to the project are insightful. Things to think about include: 1) the possibilities of cooperative learning occurring between computer and students; and 2) cooperative learning being greater with partners when working with math manipulatives than with computers.

APPENDIX E

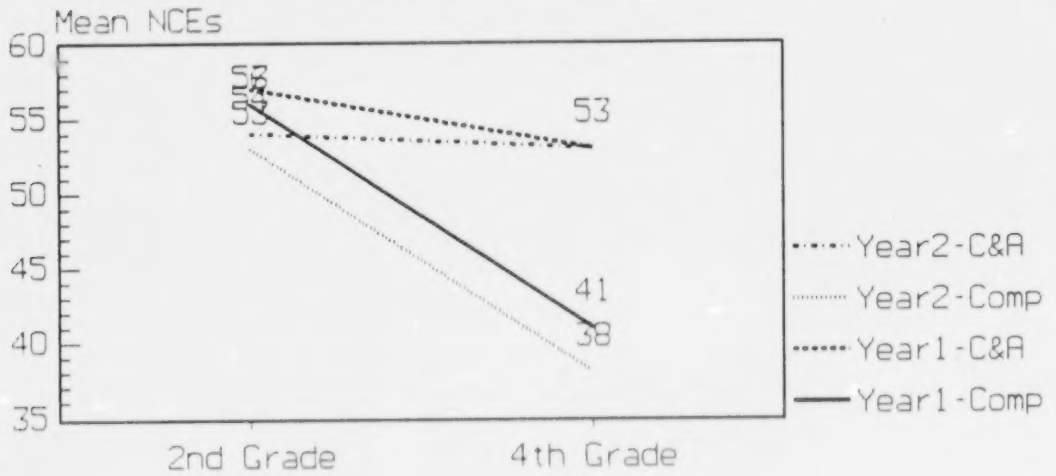
Comparison of Lighthouse Students With Previous Year's Students on CAT Math Computation - District 1 (One year of full implementation)



Comparison of Lighthouse Students With Previous Year's Students on CAT Math Concepts and Applications - District 1 (One year of full implementation)

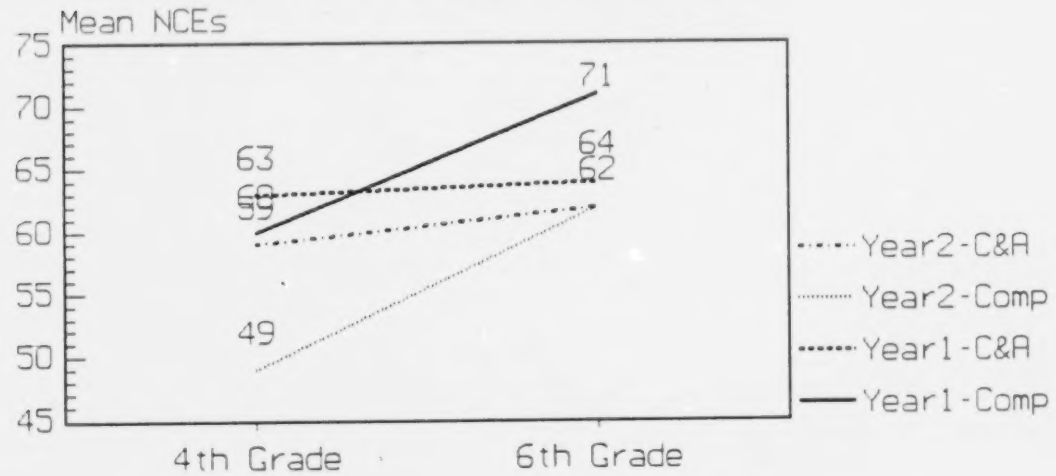


CAT Test Results for District 2
Computation vs. Concepts & Applications
Two Years of Lighthouse Implementation
Both with Third-Grade Intervention



Same students in 2nd and 4th grade

CAT Test Results for District 2
Computation vs. Concepts & Applications
Two Years of Lighthouse Implementation
Both With Fifth-Grade Intervention



Same students in 4th grade and 6th grade

CAT Test Results for District 3
Computation vs. Concepts & Applications
One versus Two Years of Lighthouse